UECM1703 Introduction of Scientific Computing

TOPIC 6 Practical

UNIVERSITI TUNKU ABDUL RAHMAN

Faculty:	FES	Unit Code:	UECM1703
Course:	AM &FM	Unit Title:	Introduction To Scientific Computing
Year:	1&2	Lecturer:	Dr Yong Chin Khian
Session:	Oct 2023		

Q1. A local Bank wants to estimate the number of foreclosures per 1000 houses sold. Bank officials think that the number of foreclosures is related to the size of the down payment made by house buyers. The following table contains foreclosure and down payment data.

Down Payment Size (% of purchase price)	Number of Foreclosures per 1000 houses
10	37
20	15
14	27
12	29
18	12
16	25

You fit the above data to $Y = \beta_0 + \beta_1 X + \epsilon$, where Y is the number of foreclosures per 1000 houses sold, and x is the size of the down payment made by house buyers. Use the stat.models module to answer the following questions:

(a)State the estmated regression function.

(b)Obtain a prediction for a new observation y_h when $x_h = 17.0$.

(c)Compute the R^2.

(d)Compute the adjusted R^2.

(e)Compute the AIC.

(f)Compute the BIC.

(g)Compute the test statistic for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$

(h)Compute the p value corresponding to the test above.

Q2. Consider the data shown below:

y	x	y	x
30037	16	13285	13
4763	10	6908	11
6908	11	30037	16
3161	9	662	6
17768	14	6908	11
662	6	6908	11
38136	17	23302	15
4763	10	3161	9

(a) Write down the python codes to fit the data to the polynomial regression model of degree 4.

(b) Continue using the codes from part (a), write down the python codes to created a scatter plot of the data and add the fitted polynomial line to the scatterplot.

Q3. You fit $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ with independent normal error terms to data in data file dfQ09c.csv. Use the stat.models package in Python to answer the following questions.

(a) State the estimated regression function.

(b) Obtain a prediction for a new observation Y_h when $x_{h_1} = 5.9$, $x_{h_2} = 30.8$, $x_{h_3} = 10.2$ and $x_{h_4} = 19.0$.

(c) Write down the python codes and the correspoding outputs that you obtained for answers in part (a) and part (b).

Q4. Consider the data below:

y	x_1	x_2	x_3	x_4
265.1	7.9	31.6	15.4	30.2
179.8	4.6	29.4	8.7	16.8
182.9	4.7	29.5	8.9	17.3
162.5	3.9	28.9	7.3	14.1
236.4	6.8	30.9	13.1	25.7
115.4	2.0	27.7	3.6	6.7
270.0	8.1	31.7	15.7	31.0
174.1	4.4	29.2	8.2	15.9
215.5	6.0	30.3	11.5	22.4
193.8	5.1	29.8	9.8	19.0

Assuming that regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$ is appropriate.

(a) Find the test statistic for testing $H_0 : \beta_4 = 0$.

(b) Find the p-value for testing $H_0 : \beta_4 = 0$.

(c) Find the test statistic for testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.

(d) Find the p-value for testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.

Q5. Consider the data shown below:

y	x	y	x
1062294	32	398299	25
164352	20	239822	22
164352	20	936299	31
86354	17	21832	12
624752	28	199420	21
10659	10	199420	21
1200597	33	718245	29
134126	19	86354	17

Fit the polynomial regression models up to order 4, then answer the following questions:

(a)Fill in the table below:

Polynomial model	Adjusted R^2	AIC	BIC
First order			
Second order			
Third order			
Fourth order			

(b)Determine a model that best fit the data based on the adjusted R^2.

(c)Write down the equation of the best fitted curve that you have selected from part (b).

(d)Created a scatter plot of the data and add the fitted polynomial line to the scatterplot.

Q6. A soft drink manufacturer use five agents (1,2,3,4,5) to handle premium distributions for its various products. The marketing director desired to study the timeliness with which the premiums are distributed. Six transactions for each agent were selected at random, and the time lapse (in days) for handling each transaction was determined. The results follow.

Agent	Time lapse (in days)					
1	23	26	29	29	27	27
2	22	23	20	27	20	24
3	20	21	29	30	10	23
4	27	34	27	28	23	28
5	31	22	28	27	32	29

Consider the model $y_{ij} = \mu_i + \epsilon_{ij}, i = 1, 2, 3, 4, 5; j = 1, 2, \ldots, 6$, where

- y_{ij} is the time lapse of the j^{th} transaction with i^{th} agent.
- μ_i is the mean time lapse of the i^{th} agent..
- $\epsilon_{ij} \sim N(0, \sigma^2)$

The hypotheses to test the equalities of means are $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ versus H_1: at least one population mean is different from the rest. Use the f_oneway package to compute the test statistic and the p value.